



(12) 发明专利

(10) 授权公告号 CN 111563442 B

(45) 授权公告日 2023.05.02

(21) 申请号 202010357986.3

G06T 7/55 (2017.01)

(22) 申请日 2020.04.29

G06T 7/10 (2017.01)

(65) 同一申请的已公布的文献号

G01S 17/86 (2020.01)

申请公布号 CN 111563442 A

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

(43) 申请公布日 2020.08.21

(56) 对比文件

(73) 专利权人 上海交通大学

CN 109816686 A, 2019.05.28

地址 200240 上海市闵行区东川路800号

CN 110097553 A, 2019.08.06

(72) 发明人 王贺升 赵小文

CN 110243370 A, 2019.09.17

(74) 专利代理机构 上海段和段律师事务所

CN 110738673 A, 2020.01.31

31334

CN 110827395 A, 2020.02.21

专利代理师 李佳俊 郭国中

CN 111080659 A, 2020.04.28

WO 2019169540 A1, 2019.09.12

(51) Int. Cl.

张国良 等.融合直接法与特征法的快速双目SLAM算法.《机器人》.2017,第39卷(第06期), 879-888.

G06V 20/10 (2022.01)

G06V 10/26 (2022.01)

G06V 10/75 (2022.01)

G06V 10/762 (2022.01)

G06V 10/764 (2022.01)

审查员 张佳期

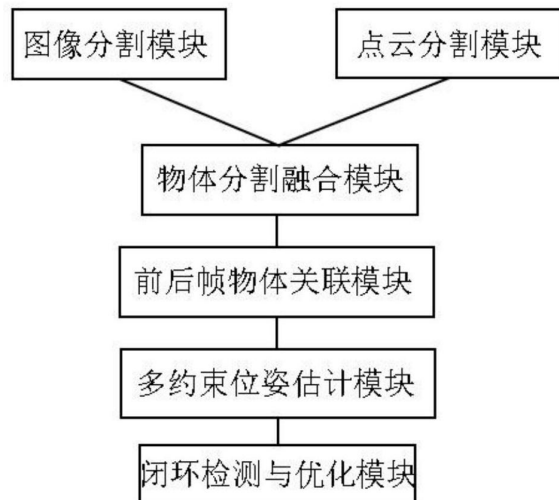
权利要求书2页 说明书8页 附图2页

(54) 发明名称

基于激光雷达的点云和相机图像数据融合的slam方法及系统

(57) 摘要

本发明提供了一种基于激光雷达的点云和相机图像数据融合的slam方法及系统,包括:提取关键帧,并对关键帧图像进行物体实例分割,得到图像中的物体实例;对关键帧的点云进行物体分割,得到点云空间中的物体对象;对图像中的物体实例和点云空间中的物体对象进行融合统一,得到物体集合;根据物体集合,对前后帧的物体进行匹配;根据前后帧的点云匹配的误差、图像的重投影误差以及特征点在前后帧中的物体类别误差,求解得到相机的位姿;根据相机的位姿将携带物体实例信息的图像注册到点云地图中,得到具有图像语义信息的点云地图。本发明提升了物体实例分割的鲁棒性,将语义约束增加到优化方程中,使得求解的位姿的精度更高。



1. 一种基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,包括:

图像分割步骤:提取关键帧,并对关键帧图像进行物体实例分割,得到图像中的物体实例;

点云分割步骤:对关键帧的点云进行物体分割,得到点云空间中的物体对象;

物体分割融合步骤:对图像中的物体实例和点云空间中的物体对象进行融合统一,得到物体集合;

前后帧物体关联步骤:根据物体集合,对前后帧的物体进行匹配;

多约束位姿估计步骤:根据前后帧的点云匹配的误差、图像的重投影误差以及特征点在前后帧中的物体类别误差,建立多约束的非线性最小二乘模型,求解得到相机的位姿;

闭环检测与优化步骤:根据相机的位姿将携带物体实例信息的图像注册到点云地图中,得到具有图像语义信息的点云地图。

2. 根据权利要求1所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,所述图像分割步骤包括:基于深度学习的图像实例分割网络Mask R-CNN来对关键帧的图像进行分割,获得图像中所有的物体实例信息;

选择在室外环境中用双目相机采集图片来作为Mask R-CNN算法的训练集,利用深度学习图像标注工具VGG Image Annotator进行标注,通过模型训练,得到训练参数;

用训练得到的权重模型对每帧关键帧进行预测,对关键帧中的物体进行目标框检测,得到剔除背景后的实例像素级掩码。

3. 根据权利要求1所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,所述点云分割步骤包括:利用欧式聚类方法完成点云的分割,通过判断两点的欧式距离是否小于聚类半径,实现激光点的分类,最终得到物体的分割结果。

4. 根据权利要求1所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,所述物体分割融合步骤包括:根据图像分割的结果和点云分割的结果确定出当前关键帧中存在的物体,剔除误检测对象。

5. 根据权利要求1所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,所述物体分割融合步骤包括:

坐标转化步骤:将图像分割得到的物体的像素点通过以下公式转化到点云空间中:

$$P' = (x' \text{ 呔 } y' \text{ 呔 } z') = RP + T$$

其中,R为从相机坐标系向激光坐标系转化的旋转矩阵,T为从相机坐标系向激光坐标系转化的平移矩阵, $P = (x, y, z)$ 为相机坐标系中的物体的点的坐标; $P' = (x', y', z')$ 为激光坐标系中点的坐标, x, y, z 分别为三维坐标系的坐标点;

分割结果融合步骤:对于转化得到的 P' 点,在点云空间中分割得到的所有物体点云中寻找近邻点 P_i ,并对每个物体转化后的点到近邻点的距离 d_i 求和,用D表示, $D = \sum_i^n d_i$,若D小于预设阈值,则判断图像分割和点云分割均得到了这一物体,并加入最终的分割结果中,否则不加入分割结果中。

6. 根据权利要求1所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,所述前后帧物体关联步骤包括:通过对物体点云提取特征并计算物体之间的特征关联距离,将前一帧中分割得到的物体和当前帧分割得到的物体进行匹配,确定对应关系;

在点云空间中得到物体包含的点,对每个物体通过特征向量进行表示,基于包围框的位置、尺寸、点的数量、点分布直方图四个特征的提取得到特征向量;

将上一帧中的每个物体与当前帧中的每个物体计算特征向量的差,作为关联距离,评估两个物体之间的相似程度,将前后帧的物体根据特征向量进行匹配,将前后帧中特征向量距离最小的物体视为同一物体,并为所有找到匹配的物体分配唯一的ID。

7.根据权利要求1所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,所述闭环检测与优化步骤包括:根据关键帧的点云特征与图像特征组成特征向量,并将当前帧的特征向量与历史帧的特征向量进行匹配完成闭环检测,构成闭环后利用点云和图像信息来优化位姿序列,并将图像信息添加到点云中构建地图。

8.根据权利要求7所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,将图像的特征点转化到点云空间中,得到其深度值和激光的强度值;

基于VLAD算法,对图像进行网格划分,对每个网格中心提取稠密sift特征、RGB颜色值,结合获取的深度值和强度值,组成特征向量;

利用k-means算法将特征描述向量进行聚类,得到预设数量的类别;

利用每个网格的特征描述向量与其聚类中心的差值,所有维度的差值组成该图像的F-VLAD向量Fusion-VLAD,将当前关键帧的F-VLAD向量与历史关键帧的F-VLAD向量进行匹配,匹配距离小于预设阈值则认为构成闭环。

9.根据权利要求8所述的基于激光雷达的点云和相机图像数据融合的slam方法,其特征在于,计算所有关键帧的点云匹配误差和图像重投影误差,相加得到联合误差的表示,并进行优化,得到优化后的位姿序列。

10.一种基于激光雷达的点云和相机图像数据融合的slam系统,其特征在于,包括:

图像分割模块:提取关键帧,并对关键帧图像进行物体实例分割,得到图像中的物体实例;

点云分割模块:对关键帧的点云进行物体分割,得到点云空间中的物体对象;

物体分割融合模块:对图像中的物体实例和点云空间中的物体对象进行融合统一,得到物体集合;

前后帧物体关联模块:根据物体集合,对前后帧的物体进行匹配;

多约束位姿估计模块:根据前后帧的点云匹配的误差、图像的重投影误差以及特征点在前后帧中的物体类别误差,建立多约束的非线性最小二乘模型,求解得到相机的位姿;

闭环检测与优化模块:根据相机的位姿将携带物体实例信息的图像注册到点云地图中,得到具有图像语义信息的点云地图。

基于激光雷达的点云和相机图像数据融合的slam方法及系统

技术领域

[0001] 本发明涉及移动机器人定位导航技术领域,具体地,涉及一种基于激光雷达的点云和相机图像数据融合的slam方法。尤其地,涉及一种移动机器人基于多传感器融合的slam方法。

背景技术

[0002] 在移动机器人定位导航领域,同步定位与建图(Simultaneous Localization and Mapping, SLAM)是指机器人在工作环境中确定自身的位姿,同时构建环境的地图。SLAM问题的本质是在环境未知的情况下机器人对环境的建模和对自身位姿的估计,为了建立周围环境的模型,需要确定自身的位姿,同时为了确定自身的位姿,又要对环境进行建模。

[0003] SLAM问题常常可以分为前端的数据关联和后端的回环优化,现有基于激光雷达和视觉的SLAM方法中,均围绕这两个部分来解决。基于视觉的方法分为直接法和特征点法,直接法将所有像素点均加入投影误差计算,而特征点法仅提取特征点来进行匹配和位姿计算,最终得到特征点地图;基于激光雷达的方法通过点云的匹配来求解位姿并将点云注册到点云地图中。传统的slam方法依赖于点、线等低层次的特征信息来估计位姿,在复杂的室外环境中,尤其是特征稀疏的场景中,其鲁棒性较低。近年来也有语义slam方法,通过语义分割来完成位姿的估计或回环检测,将slam的问题解决方法提升到语义层面。

[0004] 例如专利文献CN110781262A公开的基于视觉SLAM的语义地图的构建方法,该方法包括:通过RGB-D相机采集目标环境图像信息;根据所述目标环境图像信息选取关键帧并确定关键帧的相机位姿;对所述关键帧进行语义分割,预测图像像素的类别;利用所述关键帧的相机位姿以及所述预测的关键帧像素的类别,采用八叉树结构构建三维语义地图。

[0005] 专利文献CN110533720A(申请号:201910768052.6)公开了一种基于联合约束的语义SLAM系统与方法,旨在解决解决在像素深度值不稳定情况下相机位姿计算不准确和当动态目标占据相机视野大部分空间时无法计算相机位姿的问题,通过深度约束方法提高了相机位姿估计的准确性,采用极线约束方法提高了相机轨迹的完整性。实现方法为:数据采集模块获取图像序列;神经网络模块获取检测图像和实例分割图像;联合约束模块获取不同的特征点类别集合;数据融合模块获取静态目标实例分割图像和动态目标实例分割图像;视觉前端模块获取深度相机的位姿和三维空间中的路标点集合;后端优化模块获取全局最优的深度相机位姿和路标点;语义地图模块获取语义点云地图。

发明内容

[0006] 针对现有技术中的缺陷,本发明的目的是提供一种基于激光雷达的点云和相机图像数据融合的slam方法及系统。

[0007] 根据本发明提供的基于激光雷达的点云和相机图像数据融合的slam方法,包括:

[0008] 图像分割步骤:提取关键帧,并对关键帧图像进行物体实例分割,得到图像中的物体实例;

- [0009] 点云分割步骤:对关键帧的点云进行物体分割,得到点云空间中的物体对象;
- [0010] 物体分割融合步骤:对图像中的物体实例和点云空间中的物体对象进行融合统一,得到物体集合;
- [0011] 前后帧物体关联步骤:根据物体集合,对前后帧的物体进行匹配;
- [0012] 多约束位姿估计步骤:根据前后帧的点云匹配的误差、图像的重投影误差以及特征点在前后帧中的物体类别误差,建立多约束的非线性最小二乘模型,求解得到相机的位姿;
- [0013] 闭环检测与优化步骤:根据相机的位姿将携带物体实例信息的图像注册到点云地图中,得到具有图像语义信息的点云地图。
- [0014] 优选的,所述图像分割步骤包括:基于深度学习的图像实例分割网络Mask R-CNN来对关键帧的图像进行分割,获得图像中所有的物体实例信息;
- [0015] 选择在室外环境中用双目相机采集图片来作为Mask R-CNN算法的训练集,利用深度学习图像标注工具VGG Image Annotator进行标注,通过模型训练,得到训练参数;
- [0016] 用训练得到的权重模型对每帧关键帧进行预测,对关键帧中的物体进行目标框检测,得到剔除背景后的实例像素级掩码。
- [0017] 优选的,所述点云分割步骤包括:利用欧式聚类方法完成点云的分割,通过判断两点的欧式距离是否小于聚类半径,实现激光点的分类,最终得到物体的分割结果。
- [0018] 优选的,所述物体分割融合步骤包括:根据图像分割的结果和点云分割的结果确定出当前关键帧中存在的物体,剔除误检测对象。
- [0019] 优选的,所述物体分割融合步骤包括:
- [0020] 坐标转化步骤:将图像分割得到的物体的像素点通过以下公式转化到点云空间中:
- [0021] $P' = (x', y', z') = RP + T$
- [0022] 其中,R为从相机坐标系向激光坐标系转化的旋转矩阵,T为从相机坐标系向激光坐标系转化的平移矩阵, $P = (x, y, z)$ 为相机坐标系中的物体的点的坐标; $P' = (x', y', z')$ 为激光坐标系中点的坐标, x, y, z 分别为三维坐标系的坐标点;
- [0023] 分割结果融合步骤:对于转化得到的 P' 点,在点云空间中分割得到的所有物体点云中寻找近邻点 P_i ,并对每个物体转化后的点到近邻点的距离 d_i 求和,用D表示, $D = \sum_i^n d_i$,若D小于预设阈值,则判断图像分割和点云分割均得到了这一物体,并加入最终的分割结果中,否则不加入分割结果中。
- [0024] 优选的,所述前后帧物体关联步骤包括:通过对物体点云提取特征并计算物体之间的特征关联距离,将前一帧中分割得到的物体和当前帧分割得到的物体进行匹配,确定对应关系;
- [0025] 在点云空间中得到物体包含的点,对每个物体通过特征向量进行表示,基于包围框的位置、尺寸、点的数量、点分布直方图四个特征的提取得到特征向量;
- [0026] 将上一帧中的每个物体与当前帧中的每个物体计算特征向量的差,作为关联距离,评估两个物体之间的相似程度,将前后帧的物体根据特征向量进行匹配,将前后帧中特征向量距离最小的物体视为同一物体,并为所有找到匹配的物体分配唯一的ID。
- [0027] 优选的,所述闭环检测与优化步骤包括:根据关键帧的点云特征与图像特征组成

特征向量,并将当前帧的特征向量与历史帧的特征向量进行匹配完成闭环检测,构成闭环后利用点云和图像信息来优化位姿序列,并将图像信息添加到点云中构建地图。

[0028] 优选的,将图像的特征点转化到点云空间中,得到其深度值和激光的强度值;

[0029] 基于VLAD算法,对图像进行网格划分,对每个网格中心提取稠密sift特征、RGB颜色值,结合获取的深度值和强度值,组成特征向量;

[0030] 利用k-means算法将特征描述向量进行聚类,得到预设数量的类别;

[0031] 利用每个网格的特征描述向量与其聚类中心的差值,所有维度的差值组成该图像的F-VLAD向量Fusion-VLAD,将当前关键帧的F-VLAD向量与历史关键帧的F-VLAD向量进行匹配,匹配距离小于预设阈值则认为构成闭环。

[0032] 优选的,计算所有关键帧的点云匹配误差和图像重投影误差,相加得到联合误差的表示,并进行优化,得到优化后的位姿序列。

[0033] 根据本发明提供的基于激光雷达的点云和相机图像数据融合的slam系统,包括:

[0034] 图像分割模块:提取关键帧,并对关键帧图像进行物体实例分割,得到图像中的物体实例;

[0035] 点云分割模块:对关键帧的点云进行物体分割,得到点云空间中的物体对象;

[0036] 物体分割融合模块:对图像中的物体实例和点云空间中的物体对象进行融合统一,得到物体集合;

[0037] 前后帧物体关联模块:根据物体集合,对前后帧的物体进行匹配;

[0038] 多约束位姿估计模块:根据前后帧的点云匹配的误差、图像的重投影误差以及特征点在前后帧中的物体类别误差,建立多约束的非线性最小二乘模型,求解得到相机的位姿;

[0039] 闭环检测与优化模块:根据相机的位姿将携带物体实例信息的图像注册到点云地图中,得到具有图像语义信息的点云地图。

[0040] 与现有技术相比,本发明具有如下的有益效果:

[0041] 1、本发明将激光点云与视觉的物体分割信息进行融合,提升了物体实例分割的鲁棒性,同时将点云的匹配误差和图像的投影误差统一到同一个优化问题中,并考虑物体在前后帧的对应关系,将语义约束增加到优化方程中,使得求解的位姿的精度更高;

[0042] 2、本发明在闭环检测与优化中融合了点云与图像的信息,进一步优化了全局的位姿序列;

[0043] 3、本发明作为一种多传感器融合的slam方法,可以得到更好的位姿估计效果,以及具有图像信息的点云地图。

附图说明

[0044] 通过阅读参照以下附图对非限制性实施例所作的详细描述,本发明的其它特征、目的和优点将会变得更明显:

[0045] 图1为本发明的一种点云和图像融合的slam方法各模块示意图;

[0046] 图2为本发明的一种点云和图像融合的slam方法实施例流程示意图;

[0047] 图3为本发明的前后帧物体关联模块示意图;

[0048] 图4为本发明位姿估计方法和现有的方法对比的示意图。

具体实施方式

[0049] 下面结合具体实施例对本发明进行详细说明。以下实施例将有助于本领域的技术人员进一步理解本发明,但不以任何形式限制本发明。应当指出的是,对本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变化和改进。这些都属于本发明的保护范围。

[0050] 如图1所示,本发明提供一种点云和图像融合的slam方法,分别涉及图像分割模块、点云分割模块、物体分割融合模块、前后帧物体关联模块、多约束位姿估计模块、闭环检测与优化模块。如图2所示为本发明的实施例流程示意图。其中,步骤1涉及两个模块,图像分割模块用于获取每帧图像中的物体实例;点云分割模块用于通过聚类方法同步检测每帧中的物体点云;步骤2涉及一个模块,即物体分割融合模块,用于融合图像分割和点云分割的结果,得到统一的物体描述;步骤3涉及一个模块,即前后帧物体关联模块,用于将前后帧的物体进行匹配,并为每个物体赋予唯一的ID(如图3所示);步骤4涉及一个模块,即多约束位姿估计模块,用于将点云的匹配误差、图像特征点的重投影误差以及特征点的物体类别误差同时进行最小化,求解出相机的精确位姿(如图3所示);步骤5涉及一个模块,即闭环检测与优化模块,用于将关键帧的图像与点云添加到地图中,构建得到具有图像语义信息的点云地图。如图4,为本发明位姿估计方法和现有的方法对比的示意图。

[0051] 图像分割模块用于获取每帧图像中的物体实例,包含以下步骤:

[0052] 步骤1:选择在室外环境中用双目相机采集图片来作为Mask R-CNN算法的训练集,利用深度学习图像标注工具VGG Image Annotator (VIA) 进行标注,通过模型训练,得到参数。

[0053] 其中,采用的Mask R-CNN是一种开源的目标检测算法,其特点在于可以对目标的轮廓进行像素级分割,相比于其他的网络可以更准确地完成物体实例分割。

[0054] 步骤2:用训练得到的权重模型对每帧关键帧进行预测,对关键帧中的物体进行目标框检测,得到剔除背景后的实例像素级掩码。

[0055] 其中,关键帧的选择主要是考虑当前帧与上一帧的相对运动距离 d ,只有距离大于一定范围才视为关键帧,防止大量的重复帧出现,具体规则如下:

[0056] 1) 若 $d < d_{thre}$,则当前帧不作为关键帧;

[0057] 2) 若 $d \geq d_{thre}$,则当前帧作为关键帧;

[0058] 其中, d_{thre} 是允许的帧间最大运动距离。

[0059] 点云分割模块用于分割出每帧点云中的物体对象。具体实施方法如下:

[0060] 采用欧式聚类的方法对点云进行聚类分割,欧式聚类是一种基于欧式距离对点云中的点进行聚类的方法,聚类半径的设置与激光点到激光雷达的距离有关。采用这种可变聚类半径阈值的方法,在基于扫描线的三维点云聚类中能够取得更佳的效果。设置规则如下:聚类半径 r 的设置方法和激光点的距离 d 之间的关系为:

$$[0061] \quad r = \begin{cases} 0.5m & d < 15m \\ 1.0m & 15 < d \leq 30m \\ 1.5m & 30 < d \leq 50m \end{cases}$$

[0062] 其中, r 为聚类半径, d 为激光扫到的点到激光中心的距离。

[0063] 物体分割融合模块用于将图像分割与点云分割的结果进行融合,得到最终的物体

分割结果。包含以下步骤：

[0064] 步骤1:将图像分割得到的物体的像素点通过以下公式转化到点云空间中：

$$[0065] \quad P' = (x', y', z') = RP + T$$

[0066] 其中,R为从相机坐标系向激光坐标系转化的旋转矩阵,T为从相机坐标系向激光坐标系转化的平移矩阵,P=(x,y,z)为相机坐标系中的物体的点的坐标;P'=(x',y',z')为激光坐标系中点的坐标;

[0067] 步骤2:将分割结果进行融合。对于转化得到的P'点,在点云空间的物体点云集合中寻找近邻点P_i,并对每个物体其中的点到近邻点的距离d_i求和,用D表示, $D = \sum_i^n d_i$,假设D小于一定阈值,则认为图像分割和点云分割均得到了该物体,将其加入最终的分割结果中,否则不将其加入分割结果中。

[0068] 前后帧物体关联模块用于将前后帧的物体进行匹配,并为每个物体赋予唯一的ID,包含以下步骤:

[0069] 步骤1:物体特征提取。根据物体包含的点,对每个物体通过特征向量进行表示,基于包围框的位置、尺寸、点的数量、点分布直方图四个特征的提取得到其特征向量。

[0070] 其中,物体点云的包围框位置、尺寸、点的数量都可以简单地获得,点分布直方图是通过统计所有点在x,y,z三个方向的每个单位长度上的分布数量得到的。基于上述四种特征,可分别计算两个物体每种特征的距离,用于衡量该两个物体该特征的相似程度,具体如下所示:

$$[0071] \quad d_1 = \sqrt{(x_1^i - x_2^j)^2 + (y_1^i - y_2^j)^2}$$

$$[0072] \quad d_2 = l_1^i - l_2^j$$

$$[0073] \quad d_3 = n_1^i - n_2^j$$

$$[0074] \quad d_4 = \sqrt{\left(\sum_{t=0}^n (f_1^i[t] - f_2^j[t])^2 \right)}$$

[0075] 其中,第一个公式中(x₁ⁱ,y₁ⁱ)表示上一帧中的第i和物体的质心坐标,(x₂^j,y₂^j)表示当前帧中的第j个物体的质心坐标,其余公式中的变量分别表示尺寸、点数量和直方图特征。

[0076] 步骤2:物体关联步骤:将上一帧中的每个物体与当前帧中的每个物体计算特征向量的差,作为“关联距离”,如下式所示:

$$[0077] \quad d_{ij} = \sum_{k=1,2,3,4} d_k \cdot w_k$$

[0078] 其中,d_{ij}表示上一帧的物体i和当前帧的物体j之间的关联距离,d_k表示每种特征计算的距离,w_k表示特征的权重,由于经过仿真发现位置属性和直方图属性更能反映物体之间的相似性,而其它两条属性容易受到点云分割结果的影响,尤其在点云存在局部遮挡等情况下物体的尺寸和扫描到的点数量可能出现波动,因此在实际使用中本文给予了d₁与d₄的更大的权重。

[0079] 基于计算的关联距离,对上一帧中的每个物体,将当前帧中的与其关联距离最小

的物体视为同一物体,并为所有找到匹配的物体分配唯一的ID。

[0080] 多约束位姿估计模块用于将点云的匹配误差、图像特征点的重投影误差以及特征点的物体语义约束误差同时进行最小化,求解出相机的精确位姿。具体地,

[0081] 步骤1:计算点云的匹配误差和物体语义约束的误差

[0082] 传统的ICP算法仅仅依赖点与点之间的距离最小来作为约束条件,假设M,P是两个点集,其中P为待配准点集,而M为目标点集,取P中的每个点 P_i ,在M中找出距离 P_i 最近的 M_i ,两个点集之间的位姿变换关系为(R,T),则其最小化的目标函数为:

$$[0083] \quad E_{base} = \frac{1}{n} \sum_{i=1}^n \|R \cdot P_i + T - M_i\|_2^2$$

[0084] 与语义slam中的语义误差项一致,用下式表示物体类别的误差项:

$$[0085] \quad E_{label} = \sum_{i=1}^n w_i^{(c)} \log P(q_i | T_k, X_i, Z_i = c)$$

[0086] 其中, $w_i^{(c)}$ 表示点 P_i 属于物体C的概率, $P(q_i | T_k, X_i, Z_i = c)$ 表示在该条件下对应点 q_i 属于物体C的概率。于是,两项误差的总和项为:

$$[0087] \quad E_1 = E_{base} + \lambda E_{label}$$

[0088] 其中, λ 为权重因子,是用同一物体在前后帧的关联距离通过高斯函数计算得到,即关联距离越大,表示相似程度越低,后面的语义项权重越小。

[0089] 步骤2:计算图像特征点的重投影误差,对于两帧之间n组投影点,其重投影误差为:

$$[0090] \quad E_2 = \sum_{i=1}^n \|p_i - p'_i\|_2^2$$

[0091] 步骤3:构建多约束优化方程,求解相机的位姿:

$$[0092] \quad \{T_1\}, \{d_1\} = \arg \min (\beta E_1 + (1-\beta) \alpha E_2)$$

[0093] 其中,参数 β 为权重因子,与环境有关,在图像特征比较丰富的场景下 β 较小,参数 α 是从像素误差到空间距离误差的转换因子,和相机的分辨率以及激光投射的范围有关。

[0094] 闭环检测与优化模块用于执行闭环检测、位姿优化以及将关键帧的图像与点云添加到地图中,构建得到具有图像语义信息的点云地图。具体地:

[0095] 步骤1:将图像的sift特征点转化到点云空间中,得到其深度值和激光的强度值,基于VLAD算法,首先对图像进行网格划分,对每个网格中心提取稠密sift特征、RGB颜色值,结合获取的深度值和强度值,组成特征向量。

[0096] 步骤2:利用k-means算法将特征描述向量进行聚类,得到一定数量的类别,即一定数量的聚类中心(假设为k);利用每个网格的特征描述向量与其聚类中心的差值,所有维度的差值组成该图像的F-VLAD向量(Fusion-VLAD),表示为 $\varphi = (\psi_1, \psi_2, \psi_3, \dots, \psi_n)$;

$$[0097] \quad \psi_i = \sum_{x_j, N(x_j)=c} x_j - c$$

[0098] 其中, x_j 是该图像的一个特征点, x_j 维度为d, $N(x)$ 是离 x_j 最近的聚类中心, ψ_i 是以c

为聚类中心的特征点 x_j 的每一维的值与 c 的每一维的值的差的和。由于这种方法使得该向量的维度较高($k \times d$)，使用PCA降维法对F-VLAD向量进行处理，得到最终的向量表示。

[0099] 步骤3:将当前关键帧的F-VLAD向量与历史关键帧的F-VLAD向量计算欧式距离，距离小于一定的阈值则认为构成闭环。该阈值在离线环境下通过计算不同相似程度的图像的F-VLAD向量差可以获得。

[0100] 步骤4:检测到闭环后，利用所有关键帧点云的iCP匹配和图像的重投影误差进行联合优化，假设一共有 n 个关键帧，点云的匹配误差和图像的重投影误差分别是：

$$[0101] \quad E_p = \sum_{i=1}^n E_i$$

$$[0102] \quad E_c = \sum_{j=1}^n E_j$$

[0103] 其中， E_i, E_j 分别是在每个关键帧的误差，则优化后的位姿序列及其协方差为：

$$[0104] \quad \{\hat{T}\}, \{\hat{d}\} = \arg \min (\beta E_p + (1 - \beta) \alpha E_c)$$

[0105] 步骤5:构建地图，需要将关键帧的图像映射到三维点云中，关键帧的像素点用 $[u_j, v_j]$ 表示，其中， u_j, v_j 分别是横纵坐标，将关键帧映射到三维点云中，如下式所示：

$$[0106] \quad [x_j, y_j, z_j]^T = \exp(\hat{\xi}_i)^{-1} \cdot K^{-1} \cdot T^{-1} \cdot [u_j, v_j]^{-1}$$

[0107] 其中， $\exp(\hat{\xi}_i)^{-1}$ 是相机位姿的李群表示， K 是相机的内参矩阵， T 是相机和激光雷达之间的标定矩阵。

[0108] 根据本发明提供的基于激光雷达的点云和相机图像数据融合的slam系统，包括：

[0109] 图像分割模块：提取关键帧，并对关键帧图像进行物体实例分割，得到图像中的物体实例；

[0110] 点云分割模块：对关键帧的点云进行物体分割，得到点云空间中的物体对象；

[0111] 物体分割融合模块：对图像中的物体实例和点云空间中的物体对象进行融合统一，得到物体集合；

[0112] 前后帧物体关联模块：根据物体集合，对前后帧的物体进行匹配；

[0113] 多约束位姿估计模块：根据前后帧的点云匹配的误差、图像的重投影误差以及特征点在前后帧中的物体类别误差，建立多约束的非线性最小二乘模型，求解得到相机的位姿；

[0114] 闭环检测与优化模块：根据相机的位姿将携带物体实例信息的图像注册到点云地图中，得到具有图像语义信息的点云地图。

[0115] 本领域技术人员知道，除了以纯计算机可读程序代码方式实现本发明提供的系统、装置及其各个模块以外，完全可以通过将方法步骤进行逻辑编程来使得本发明提供的系统、装置及其各个模块以逻辑门、开关、专用集成电路、可编程逻辑控制器以及嵌入式微控制器等的形式来实现相同程序。所以，本发明提供的系统、装置及其各个模块可以被认为是一种硬件部件，而对其内包括的用于实现各种程序的模块也可以视为硬件部件内的结构；也可以将用于实现各种功能的模块视为既可以是实现方法的软件程序又可以是硬件部

件内的结构。

[0116] 以上对本发明的具体实施例进行了描述。需要理解的是,本发明并不局限于上述特定实施方式,本领域技术人员可以在权利要求的范围内做出各种变化或修改,这并不影响本发明的实质内容。在不冲突的情况下,本申请的实施例和实施例中的特征可以任意相互组合。

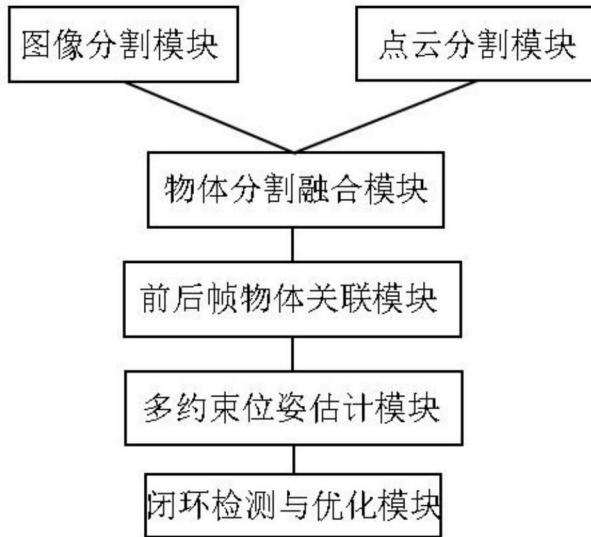


图1

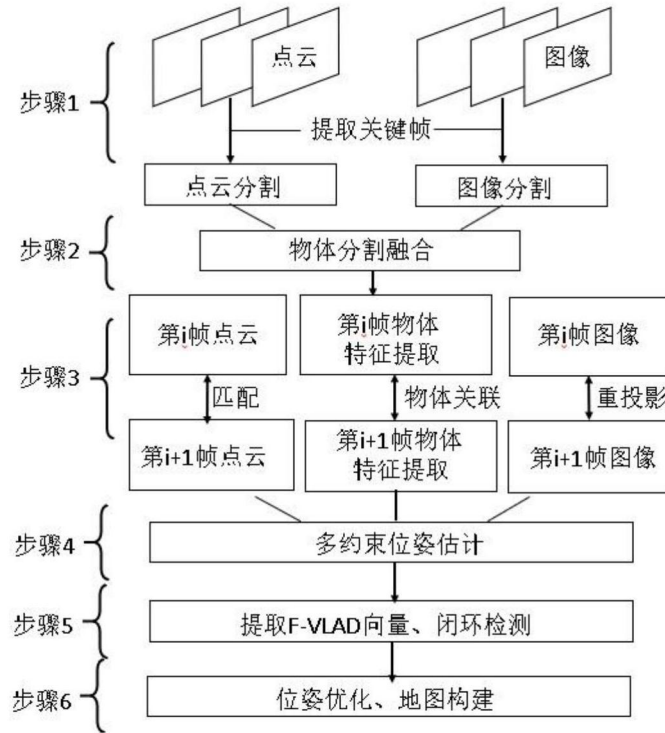


图2

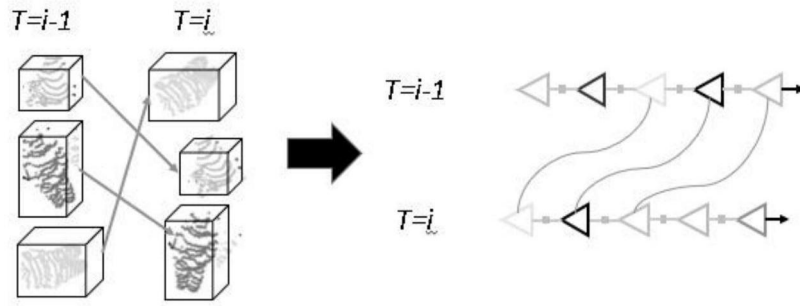


图3

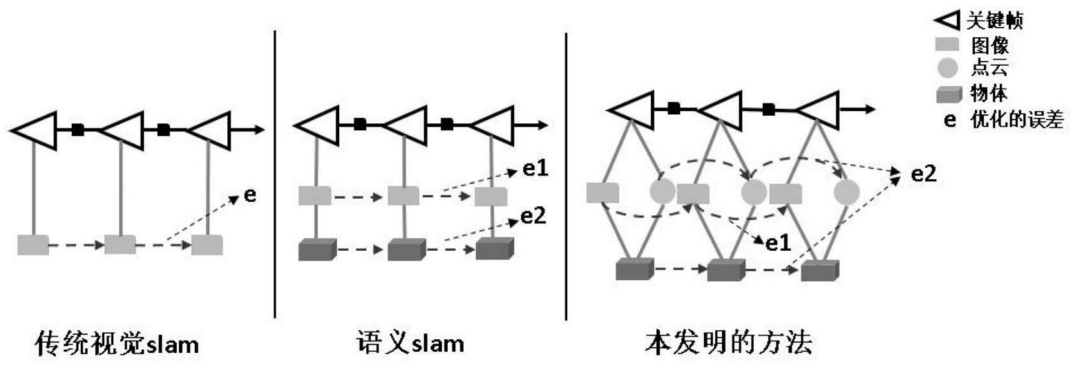


图4